Degree distribution of complex networks from statistical mechanics principle

Ginestra Bianconi

The Abdus Salam International Center for Theoretical Physics,

Strada Costiera 11, 34014 Trieste, Italy

Abstract

In this paper we describe the emergence of scale-free degree distributions from statistical mechanics principles. We define an energy associated to a degree sequence as the logarithm of the number of indistinguishable simple networks it is possible to draw given the degree sequence. Keeping fixed the total number of nodes and links, we show that the energy of scale-free distribution is much higher than the energy associated to the degree sequence of regular random graphs. This results unable us to estimate the annealed average of the number of distinguishable simple graphs it is possible to draw given a scale-free distribution with structural cutoff. In particular we shaw that this number for large networks is strongly suppressed for power -law exponent $\gamma \to 2$.

PACS numbers: :89.75.Hc,89.75.Fb,89.75.Da

Recently a large variety of complex systems, from the Internet to the protein interactions in the cell, have been described in terms of the underlining complex networks [1, 2, 3]. In many cases the topology of these networks is the outcome of a self-organized stochastic process since there is no an a priori design of the connections. Nevertheless these structures evolve in order to perform some special task. Therefore it is crucial to ask how much complex networks are far from optimal performance. To answer this question the optimality of the networks has been defined with respect to a variety of different specific prerequisites [4, 5, 6, 7, 8, 9]: *i)* respect to a specific function [4, 5, 6], or *ii)* to their dynamics [7, 8], or *iii)* to some topological robustness features [9]. Nevertheless, many real networks show some universality in their topology [1, 2]. A major universal character of these networks is given by their scale-free degree distribution which is a general property of a vast class of complex systems [10]. In fact, a large variety of real networks are scale-free with diverging second moment of their degree distribution $\langle k^2 \rangle$ while others are finite-scale, with a finite second moment $\langle k^2 \rangle$. These two general network distributions have very strong impact also on the dynamics defined on these graphs as it has been widely discussed in the literature [11]. Until now the way to explain the appearance of the recently identified scale-free structure is two-fold: on one side there are growing network models which assume that scale-free networks dynamically emerge from a growing process in which new links are added following the preferential attachment rule [1, 2, 3], on the other side there are equilibrium networks which by means of some externally imposed hidden variable present on each node show up a scale-free degree distribution [12, 13]. We note here that because of the so widely observed degree distributions of the networks, further attention must be put on their details: i.e. on the deviations from random scale-free or finite-scale distributions which makes them specific to their function as it has been indicated for example by works on motifs and community identification [14, 15].

In this paper we propose a method which describes the emergence of scale-free networks from statistical mechanics principles. We limit ourself to the case in which the embedding space, if it exists, plays a limited role in the network wiring and we allow any two nodes of the network to be linked. We take a very simple and general assumption: we assume that the degree distribution of the network minimize a partition function associated to the network. A fundamental role in this partition function is played by the energy associated to the degree distribution. We define this energy as the logarithm of the total number $\mathcal{N}_G$ of

indistinguishable simple networks it is possible to draw by wiring the edges given a certain degree distribution. Consequently, the energy is a measure of the redundancy present in the space of allowed simple networks given a degree distribution. This space is not an absolute abstract space but it is a well known space considered in many applications. Indeed for example it is sampled by the widely used randomization algorithm [16], by swapping pairs of links without changing the degree distribution. The formulation of our problem assumes the aspect of some type of statistical mechanics of the networks, a direction which has been pursued also by other authors [17, 18, 19, 20, 21, 22]. However here we don't fix an a priori degree distribution, or a preferential attachment dynamics like in Ref. [17, 20, 21, 22] but we take an unsupervised approach in which we derive the more probable distribution of a graph of N nodes and L links, which minimize a free energy defined in terms of a single external parameter z. On the other side the free energy differs from the free energy of [18, 19] because the partition function is defined directly over the degree distribution. In this work we show that scale-free degree distributions with different exponent $\gamma > 2$ minimize the free energy of our problem. The energy of such graphs is an extensive quantity which decrease with γ showing that networks with $\gamma \to \infty$ which are regular random networks with degree distribution infinitevely peacked around their average value, have much lower energy than networks with lower γ and diverging second moment of the connectivity $< k^2 >$. These founding unable us to draw some conclusion on the nature of the space of simple graphs associated with given degree sequence. Making the estimation of the average number of simple graphs $\mathcal{N}_{SG}$ and showing that in the thermodynamic limit and when a structural cutoff is considered this number is strongly suppressed as $\gamma \to 2$.

In order to define our partition function let's consider a network with N nodes and $2L$ edges with degree sequence $\{k_1, \dots, k_N\}$ associated to the degree distribution N_k. We define the energy $E(\{N_k\})$ associated to the degree distribution of the network as the logarithm of the number $\mathcal{N}_G$ of indistinguishable simple networks it is possible to draw given the degree sequence, i.e.

$$E(\{N_k\}) = \log(\mathcal{N}_G). \tag{1}$$

where a simple network is a graph without tadpoles and double links. The number $\mathcal{N}_G$ can be expressed as

$$\mathcal{N}_G = e^{E(\{N_k\})} = \prod_k k!^{N_k}. \tag{2}$$

In fact, for every simple graph associated with the given distribution every permutation of the edges departing from each node generates the same set of links. These permutations are given by $\prod_k k!^{N_k}$, and consequently we derive Eq. (2).

On the other side, the number of ways $\mathcal{N}_{\{N_k\}}$ in which we can distribute $2L$ edges into any degree sequence $\{k_1, \dots, k_N\}$ of distribution $\{N_k\}$ is given by

$$\mathcal{N}_{\{N_k\}} = \frac{(2L)!}{\prod_k (kN_k)!}. \tag{3}$$

where we consider unlabeled nodes. Thus we can define an entropy $S(\{N_k\})$ of each distribution as

$$e^{S(\{N_k\})} = \mathcal{N}_{\{N_k\}} = \frac{(2L)!}{\prod_k (kN_k)!}. \tag{4}$$

Once we have defined the energy $E(\{N_k\})$ and the entropy $S(\{N_k\})$ associated to a network degree distribution, we can define a partition function of the network.

Proceeding as in standard statistical mechanics, we define a normalized partition function Z of the network as the sum over all microstates of the problem, i.e. degree distributions, with given energy $E(\{N_k\})$ and entropy $S(\{N_k\})$

$$Z = \frac{1}{(2L)!} \sum_{\{N_k\}}{}' e^{-\frac{1}{z}E(\{N_k\})+S(\{N_k\})}. \tag{5}$$

In other words we would like to know which are the more likely distributions $\{N_k\}$ which minimize the free energy of the network $F(\{N_k\}) = E(\{N_k\}) - zS(\{N_k\})$. The role of the parameter z is to measure a tradeoff between the 'energetic' and the 'entropic' term in the definition of the free energy, as well as the temperature T in classical statistical mechanics. In equation (5) the sum $\sum'$ over the $\{N_k\}$ distributions is extended only to $\{N_k\}$ for which the total number of nodes N and the total number of links L in the network is fixed, i.e.

$$\begin{aligned} \sum_k N_k &= N \\ \sum_k kN_k &= 2L. \end{aligned} \tag{6}$$

To enforce these conditions we introduce in (5) the delta functions in the integral form

$$\delta(\sum_k N_k - N)\delta(\sum_k kN_k - 2L) = \int_0^{2\pi} \frac{d\lambda}{2\pi} \int_0^{2\pi} \frac{d\nu}{2\pi} e^{i\lambda(\sum_k kN_k - 2L) + i\nu(\sum_k N_k - N)}. \tag{7}$$

Performing

$$
\begin{aligned}
Z &= \frac{1}{(2L)!}\int\frac{d\lambda}{2\pi}\int\frac{d\nu}{2\pi}\sum_{\{N_k\}}\exp\left[-\frac{1}{z}E(\{N_k\})+S(\{N_k\})-i\lambda(2L-\sum_k kN_k)-i\nu(N-\sum_k N_k)\right].\\
&= \int\frac{d\lambda}{2\pi}\int\frac{d\nu}{2\pi}\exp\left[-i\lambda 2L-i\nu N+\sum_k\log G_k(\lambda,\nu)\right]=\int\frac{d\lambda}{2\pi}\int\frac{d\nu}{2\pi}\exp[Nf(\lambda,\nu)] \qquad (8)
\end{aligned}
$$

where

$$
G_k(\lambda,\nu)=\sum_{N_k}\frac{1}{(kN_k)!}\exp\left\{N_k\left[i\lambda k+i\nu-\frac{1}{z}\log(k!)\right]\right\}. \qquad (9)
$$

Assuming that the sum over all N_k can be approximated by the sum over all $L_k = kN_k = 1,2,\ldots\infty$ we get $\log\left[G_k(\lambda,\nu)\right]=\exp\left[i\lambda+i\nu/k-\frac{1}{kz}\log(k!)\right]$ and

$$
f(\lambda,\nu)=-i\langle k\rangle\lambda-i\nu+\frac{1}{N}\sum_k e^{i\lambda+i\nu/k-\frac{1}{zk}\log(k!)}, \qquad (10)
$$

where $<k>=2L/N$ indicates the average degree of the network. By evaluating (8) at the saddle point, deriving the argument of the exponential respect to λ and ν, we obtain

$$
\begin{aligned}
\langle k\rangle &= \frac{1}{N}\sum_k e^{i\lambda+i\nu/k-\frac{1}{zk}\log(k!)}\\
1 &= \frac{1}{N}\sum_k\frac{1}{k}e^{i\lambda+i\nu/k-\frac{1}{zk}\log(k!)} \qquad (11)
\end{aligned}
$$

and the marginal probability that $L_k = kn = \ell$ is given by

$$
P(L_k=\ell=nk)=\frac{1}{\ell!}e^{-\ell/k\log(k!)}\frac{Z_k(L,\ell,N)}{Z(L)}, \qquad (12)
$$

with

$$
Z_k(L,\ell,N)=\int\frac{d\lambda}{2\pi}\int\frac{d\nu}{2\pi}\exp[Nf_k(\lambda,\nu,\ell)] \qquad (13)
$$

and $f_k(\lambda,\nu,\ell)=i(\langle k\rangle-\ell/N)\lambda-i\nu(1-\ell/(kN))+\frac{1}{N}\sum_{s\neq k}\exp[i\lambda+i\nu/s-\frac{1}{zs}\log(s!)]$. If we develop (12) for $\ell\ll L$ and we use the Stirling approximation for factorials, we get that each variable L_k is a Poisson variable with mean $<L_k>$ satisfying

$$
\frac{<L_k>}{k}=<N_k>=k^{-\frac{1}{z}-1}e^{\lambda+\frac{1}{z}}e^{\nu/k} \qquad (14)
$$

which describe the optimal degree distribution for our problem. If we restrict ourself to the networks with finite average degree in the thermodynamic limit, the allowed values of z are $z<1$. From the expression (14) of the optimal degree distribution, if $z\in(0,1)$ the

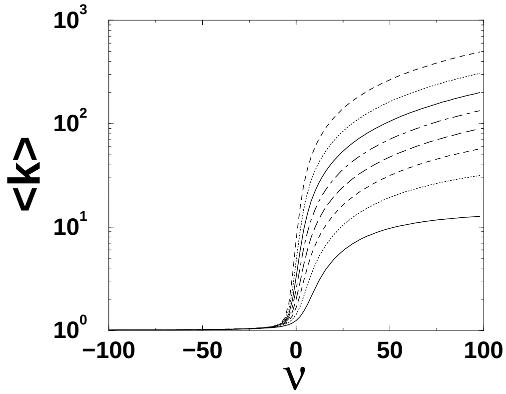


FIG. 1: Average degree $\langle k\rangle$ of the optimal distribution of as a function of ν for networks with $z = 0.2$ (solid line) $z = 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ and $z = 0.9$ (dashed line).

optimal degree distribution is scale-free with a power-law tail characterized by the exponent $\gamma = \frac{1}{z}+1$. In this case the distribution (14) always diverges in zero, thus we necessarily must impose that there are no isolated nodes in the network. The parameter $\nu \neq 0$ modulates the average degree of the graph constituting for $\nu > 0$ an effective lower cutoff of the distribution wherever the upper cutoff K of the degrees is the natural cutoff of the distribution (14). A different scenario arises if $z < 0$, when the optimal network (14) has a power-law degree distribution increasing with the degree k. In this case the Lyapunov functions ν and λ cannot fix the average degree unless one introduces by hand an upper cutoff K in the degree of the nodes of the order of magnitude of the average degree $\langle k\rangle$. We note here that also for $z \in (0, 1)$ it could be convenient to set by hand a structural cutoff $K \sim N^{1/2}$ for $z > 1/2$ in order to obtain an uncorrelated network. In the following we will consider only the power-law case $z \in (0, 1)$.

From equations (11) we derive for the Lagrangian multipliers λ and ν in $z \in (0, 1)$:

$$\begin{aligned}
\nu^{-\frac{1}{z}} e^{\lambda+\frac{1}{z}} \left[\Gamma(\frac{1}{z},\frac{\nu}{K}) - \Gamma(\frac{1}{z},\nu)\right] &= N \\
\nu\left[\Gamma\left(\frac{1}{z}-1,\frac{\nu}{K}\right) - \Gamma\left(\frac{1}{z}-1,\nu\right)\right] &= \langle k\rangle \left[\Gamma\left(\frac{1}{z},\frac{\nu}{K}\right) - \Gamma\left(\frac{1}{z},\nu\right)\right] \\
\nu^{-\frac{1}{z}} e^{\lambda+\frac{1}{z}} \left[\Gamma(\frac{1}{z}) - \Gamma(\frac{1}{z},\frac{\nu}{K})\right] &= 1,
\end{aligned} \qquad (15)$$

where $\Gamma(a, b)$ indicates the incomplete Gamma function. This system of equations is solvable provided $\langle k\rangle > 1$, as can be seen from Figure 1 where we plot the average degree of a network

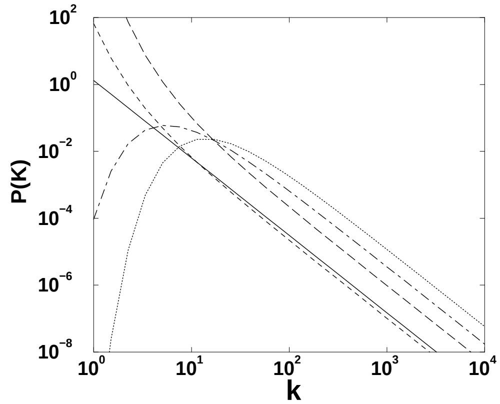


FIG. 2: The normalized degree distribution $P(k)$ for $z = 0.75$ as a function of the average degree of the network. Data are shown for $\langle k \rangle = 1.3, 2, 4, 40, 100$.

as a function of ν for different values of z. In fact, for $z \in (0, 1)$ there are highly connected nodes in the network and the degree distribution for large degrees k decays as a power-law with an exponent γ fixed by the value of z. On the other side, for small value of the degrees k the degree distribution deviates from the simple power-law and N_k depends strongly on the value of the Lagrangian multiplier ν. For low average connectivities,i.e.

$$\langle k \rangle = \langle k \rangle_0 < \frac{1}{1-z} \tag{16}$$

the solution of Eqs. (15) involves a negative value of ν. Accordingly, low degree nodes are more frequent than expected by a simple power-law while for $\langle k \rangle > \langle k \rangle_0$, $\nu > 0$ and the low connected nodes are less probable than predicted by a simple power-law.

In Figure 2 we show the optimal distributions which solve these equations at different values of the average degree $\langle k \rangle$.

Given the optimal distributions (14) we can calculate the energy of the network as a function of z at fixed average degree $\langle k \rangle$. In Figure 3 we present the energy of the optimal graph as a function of z for different average connectivities.

The energy has a minimum in the limit $z \to 0$ when the optimal degree distribution is infinitely peaked around the average, and thus the graph is a random regular graph. On the contrary, in the limit $z \to 1$ where the degree distribution has a power-law exponent $\gamma \to 2$ the energy $E(\{N_k\})$ is at the maximum.

From the derivation of our model it is evident the connection with "ball in the box"

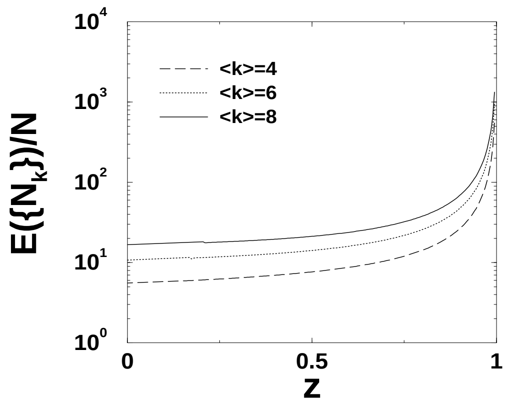


FIG. 3: Energy of the network as a function of the parameter z, for an average degree $\langle k \rangle = 4, 6, 8$ in the limit $N \to \infty$.

problems [22, 23, 24]. In our approach the "boxes" map to the degree of the nodes and the "balls" map to the edges of the graph. This makes a crucial difference respect with the model [22], in which the "boxes" map to the nodes of the graph. Although their formal similarities this difference make the two model very different in their conclusions.

We would like to indicate here that if we introduce by hand a structural cutoff $K \sim N^{1/2}$ we can assume that the networks described in this paper are randomly wired. In this case we can evaluate the number $\mathcal{N}_{SG}$ of distinguishable simple graphs is it possible to construct given the degree distribution (14). This number is approximated by

$$\mathcal{N}_{SG} \propto \frac{(2L)!! e^{-\frac{1}{2}\left(\frac{<k^2>}{<k>}\right)^2}}{e^{E(\{N_k\})}} \tag{17}$$

In fact the total number of wiring it is possible to draw given $2L$ edges is given by $(2L)!!$. This number include all type of possible wiring of the edges including the ones which give rise to graphs which are not simple. Assuming that the graph is randomly wired, i.e. that the probability that a node with k_i edges connect to a node with k_j edges is a Poisson variable with average $k_i k_j/(<k> N)$ the probability Π that the graph is simple is equal to [25]

$$\Pi = \prod_{i,j} \left(1 + \frac{k_i k_j}{<k> N}\right) e^{-k_i k_j <k> N} \sim e^{-\frac{1}{2}\left(\frac{<k^2>}{<k>}\right)^2}. \tag{18}$$

Finally in the expression (17) for $\mathcal{N}_{SG}$ there is an additional terms which takes into account the equivalent wiring of the edges which is given by $e^{-E(\{N_k\})}$. The term $\frac{<k^2>}{<k>}$ for scale-free

graphs with cutoff $K \propto N^{1/2}$ is subleading respect to the energetic term $E(\{N_k\})$ which dominates for large network sizes N.

Consequently the total number of distinguishable simple graphs $\mathcal{N}_{SG}$ in an annealed approximation, is then a decreasing function of γ suggesting that in random scale free graphs the space of distinguishable simple random graphs is strongly suppressed as $\gamma \to 2$.

In conclusion, the statistical mechanics treatment of complex networks shown in this paper is able to put in a similar context, the emergence of scale-free networks and finite-scale networks. Scale-free degree distribution correspond to higher energy state of the network respect to finite-scale networks. Especially regular random graphs have minimal energy. Consequently the large variety of real complex networks in many technological, biological and social systems which show a scale-free degree distribution reveal a tendency have a degree distribution which maximize the energy. Furthermore we give an estimation of the annealed average size of the space of simple graphs $\mathcal{N}_{SG}$ for random scale-free networks showing that $\mathcal{N}_{SG}$ is strongly reduced as $\gamma \to 2$.

---

[1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).

[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press,Oxford,2003).

[3] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Cambridge University Press,Cambridge,2004).

[4] J. R. Banavar, A. Maritan and A. Rinaldo, *Nature* **399**, 140 (1999).

[5] M. Barthélemy and A. Flammini, cond-mat/0602081

[6] M. T. Gastner and M. E. J. Newman, cond-mat/0603278.

[7] R. Guimerá, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales and A. Arenas, *Phys. Rev. Lett.* **89**, 248701 (2002).

[8] L. Donetti, P. I. Hurtado and M. A. Muñoz, *Phys. Rev. Lett.* **95**, 188701 (2005).

[9] A. X. C. N. Valente, A. Sarkar and H. A. Stone, *Phys. Rev. Lett.* **92**, 118702 (2004).

[10] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).

[11] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000); R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).

[12] G Caldarelli, A. Capocci, P. De Los Rios and M. A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).

[13] K.-I. Goh, B. Kahng and D. Kim, *Phys. Rev. Lett.* **87**, 278701 (2001).

[14] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Science*, **298**, 824 (2002).

[15] L. Danon, J. Duch, A. Diaz-Guilera and A. Arenas, *JSTAT*, P09008 (2005).

[16] K. Sneppen and S. Maslov, *Science* **296**, 910 (2002).

[17] Z. Burda, J. D. Correia and A. Krzywicki, *Phys. Rev. E* **64**, 046118 (2001).

[18] J. Berg and M. Lässig, *Phys. Rev. Lett.* **89**, 228701 (2002).

[19] G. Palla, I. Derényi, I. Farkas and T. Vicsek, *PRE* **69**, 046117 (2004).

[20] S. N. Dorogovtsev, J. F. F. Mendes and A. Samukhin, *Nucl. Phys. B* **666** , 396 (2003).

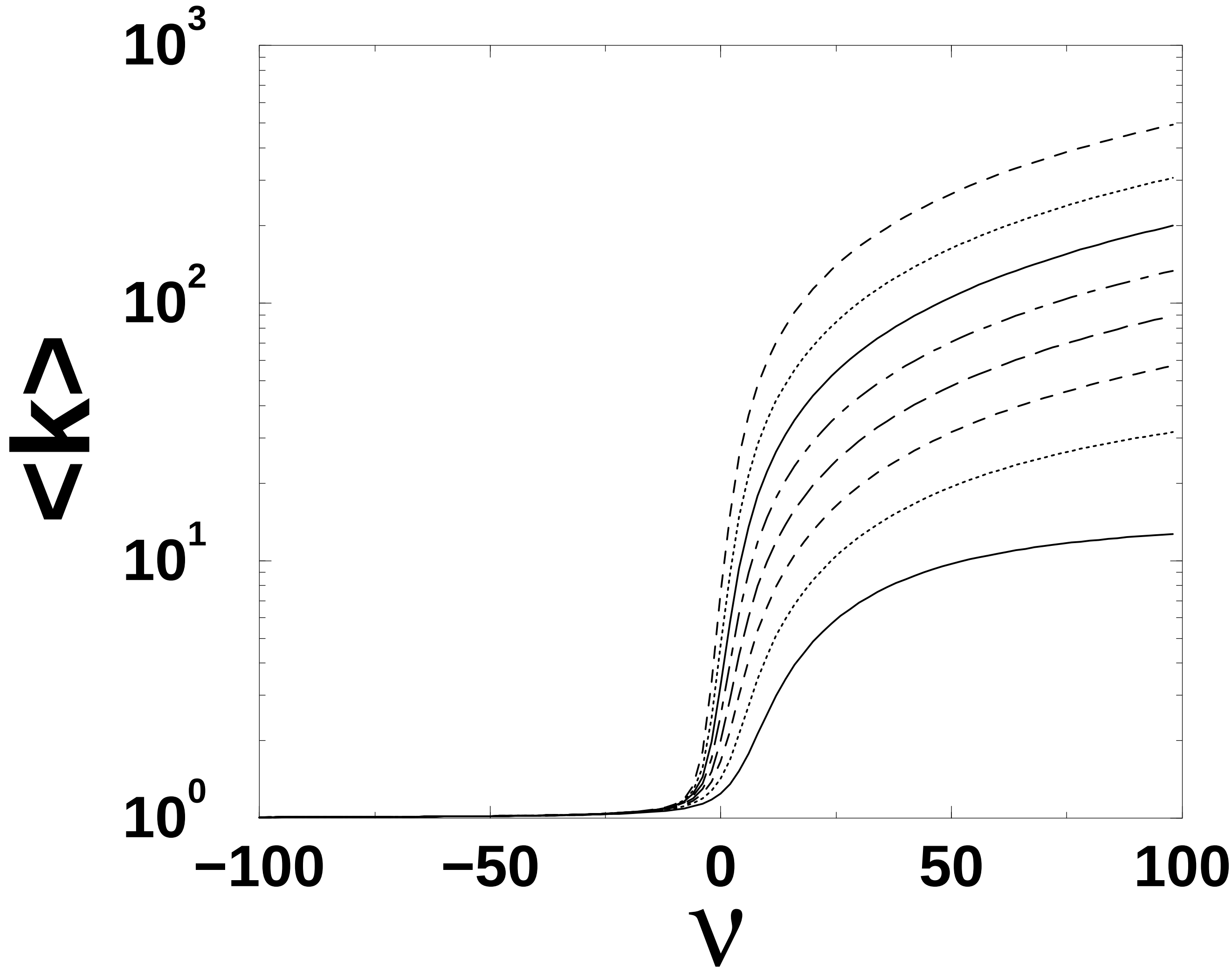
[21] J. Park and M. E. J. Newman, *Phys. Rev. E* **70**, 066117 (2004).

[22] J. Ohkubo, M. Yasuda and K. Tanaka, *Phys. Rev. E* **72**, 065104(R) (2005).

[23] F. Ritort, *Phys. Rev. Lett.* **75** 1190 (1995).

[24] P. Bialas, Z. Burda and D. Johnston, *Nucl. Phys. B* **505** (1997).

[25] G. Bianconi and M. Marsili, JSTAT P06005 (2005).



<k>
10^3
10^2
10^1
10^0
−100
−50
0
50
100
ν